OM WANI

(+91) 87-8899-1862 | Pune, IN | omwani03@gmail.com | linkedin.com/in/omwani03 | github.com/om-wani

Overview

AI Engineer with hands-on experience building production RAG pipelines, LLM-powered applications, and agentic workflows and large language models (LLMs) using LangChain, ChromaDB, FastAPI, and prompt engineering. Delivered end-to-end AI systems independently and within teams. Active FOSS contributor and a fast learner who ships.

Technical Skills

Languages: Python, JavaScript, C/C++, C#, SQL, GoLang, Bash, HTML/CSS
AI/ML & LLM: LangChain, LangGraph, LlamaIndex, RAG(Retrieval-Augmented Generation),LLMs(Large-Language Models), Agentic AI, Prompt Engineering, PyTorch, Scikit-learn, TensorFlow, OpenAI API, Claude API, Gemini API, Ollama, LM Studio, CrewAI, AutoGen, Fine-tuning, Embeddings, Inference Optimization, Transformer Models, MLOps, Natural Language Processing (NLP), Deep Learning
Vector Stores & Embeddings: ChromaDB, Pinecone, HuggingFace Sentence Transformers, LangSmith
Frameworks & APIs: FastAPI, Flask, Django, ASP.NET Web API, Node.js, React, Streamlit, REST APIs
Databases: SQL Server, MySQL, SQLite, PostgreSQL
DevOps & Tools: Git, Docker, Linux, AWS, CI/CD, VS Code, PyCharm, Model Context Protocol (MCP), n8n, Hugging Face Hub, Weights & Biases
Data & Misc: NumPy, Pandas, JSON, XML

Experience

AIML Engineer - Freelancer Aug 2025 – Present
Upperture Interactive(Om Wani) *Pune, IN*

- Architected and deployed a **custom RAG pipeline** from scratch, enabling Sales, Marketing, Recruitment, and Operations teams to query proprietary company knowledge via natural language – cutting manual information retrieval time by an estimated **60%**.
- Engineered a **production-grade customer-facing AI chatbot** that handles general inquiries and support queries end-to-end, reducing human support load and improving first-response resolution rates for **100+ daily customer interactions**.
- Delivered both systems within a **6-month engagement**, independently owning the full lifecycle: requirements gathering, architecture, development, deployment, and stakeholder demos.

Software Development Intern Feb 2025 – Jul 2025
Jade Global Software Pvt. Ltd. *Pune, IN*

- **Led a 4-person intern team** to architect and ship a proprietary **network asset scanning tool**, enabling real-time monitoring of corporate infrastructure, proactive CVE/vulnerability detection, and hardware compliance enforcement across the organization.
- Delivered the tool **under senior engineering mentorship**, applying skills across a multi-language stack (Python, Java, ASP.NET, MySQL) and gaining hands-on exposure to **industrial-scale software architecture and enterprise security protocols**.

Web Development Intern Jul 2024 – Dec 2024
BioClutch Scientific Pvt. Ltd. *Pune, IN*

- Contributed to development of a **hospital and clinic ERP suite** covering patient data management, patient knowledge base and CRM workflows, now deployed across **50+ healthcare institutions**.
- Participated in developing **embedded operating software** for new medical instruments and products being brought to market by the company.

Projects

Narayan | *Python, FastAPI, React, ChromaDB, RAG, Ollama, Langchain* Dec 2025 – Jan 2026

- Link: https://github.com/om-wani/Narayan
- Built a **full-stack RAG system** over medical research PDFs with a FastAPI backend and React/Vite frontend, delivering citation-grounded Q&A with **low retrieval latency** via production-style REST APIs.

- Engineered an **end-to-end retrieval pipeline**: PyMuPDF extraction optimized for two-column academic papers, recursive chunking with overlap, HuggingFace sentence-transformer embeddings, vector embeddings, ChromaDB vector search, and reranking, improving answer relevance by eliminating off-topic retrievals before LLM generation.
- Implemented **hallucination mitigation**: source-only answering constraints, numbered evidence snippets, inline citations, and structured metadata (source file, page, similarity score, token usage) – achieving **fully traceable, auditable** responses critical for medical use cases.
- Delivered **production-grade document lifecycle management**: content-hash deduplication, per-document scoped querying, validated upload/delete/list endpoints, and a drag-and-drop frontend with expandable evidence cards.

Fixed Asset Transfer System | *ASP.NET Web API, C#, SQL Server, JWT Auth* Apr 2025 – May 2025

- Link: https://github.com/om-wani/FATP
- Built a **role-based enterprise asset management platform** handling cross-department and cross-entity fixed asset transfers and disposals, with **dynamic multi-stakeholder approval workflows** (managers, accountants, finance controllers).
- Engineered **conditional approval routing logic** triggered by asset type and financial thresholds, eliminating manual coordination overhead across **3+ approval tiers**.
- Designed **normalized SQL Server schemas** for asset tracking, accounting records, and tamper-evident audit trails via Entity Framework Core; secured with **JWT authentication**, file upload handling, and automated PDF generation for finalized forms.

Syndicate Events | *Python, Django, GeoDjango, SQLite, Django Templates* Nov 2024 – Dec 2024

- Link: https://github.com/om-wani/SyndicateEvents
- Developed a **full-stack event aggregation platform** in Django where hosts list local events and users discover them through **geolocation-based proximity filtering** powered by GeoDjango and OpenStreetMap.
- Architected the complete stack solo: **normalized SQLite schemas** for events, accounts, saved events, and notifications; multi-parameter filtering by event type, category, and status; and a full **user auth system** (signup, login, session management) via Django's auth framework.

Education

Dr. D. Y. Patil Arts, Commerce and Science College Pune, MH, IN

Bachelor of Science in Computer Science – 7.46 CGPA *2022 – 2025*

Certifications & Achievements

IBM RAG & Agentic AI Specialization Mar 2026

Coursera

- Covered Agentic Systems, AI Security, Vector Embeddings, Generative Model Architectures, LLM Application Development, Multimodal Prompting, Responsible AI, RAG, and Tool Calling.
- Hands-on with LangChain, LangGraph, OpenAI API, Prompt Engineering, AI Orchestration, and Agentic Workflows.

Other Experiences

Pune FOSS Users' Group 2019 – Present

Volunteer & Advocate

- Promoted open-source software to **hundreds of individuals annually**; hosted a FOSS workshop at college reaching **200+ students**.

Debian Project Sep 2023 – Present

Contributor – Internationalization & Localization

- Translated parts of the **Debian Installer to Marathi**; contributed to the i18n/l10n team and attended and helped in the operations of **DebConf23 Kochi, India**.

Google Developer Student Club (GDSC-DYPACS) Aug 2022 – Dec 2024

Management Team

- Organised and hosted **workshops, tech talks, hackathons, and events** covering Google Cloud, Kubernetes, and open-source technologies; mentored **fellow students during placement season**.